Am. J. Hum. Genet. 56:461–474, 1995

Population Genetics of Dinucleotide $(dC\text{-}dA)_n \cdot (dG\text{-}dT)_n$ Polymorphisms in World Populations

Ranjan Deka,[1] Li Jin,[3,*] Mark D. Shriver,[1] Ling M. Yu,[1] Susan DeCroo,[1] Joachim Hundrieser,[4] Clareann H. Bunker,[2] Robert E. Ferrell,[1] and Ranajit Chakraborty[3]

Departments of [1]Human Genetics and [2]Epidemiology, University of Pittsburgh, Pittsburgh; [3]Genetics Centers, The University of Texas Health Science Center, Houston; and [4]Medizinische Hochschule Hannover, Hannover

Summary

We have characterized eight dinucleotide $(dC\text{-}dA)_n \cdot (dG\text{-}dT)_n$ repeat loci located on human chromosome 13q in eight human populations and in a sample of chimpanzees. Even though there is substantial variation in allele frequencies at each locus, at a given locus the most frequent alleles are shared by all human populations. The level of heterozygosity is reduced in isolated or small populations, such as the Pehuenche Indians of Chile, the Dogrib of Canada, and the New Guinea highlanders. On the other hand, larger average heterozygosities are observed in large and cosmopolitan populations, such as the Sokoto population from Nigeria and German Caucasians. Conformity with Hardy-Weinberg equilibrium is generally observed at these loci, unless (*a*) a population is isolated or small or (*b*) the repeat motif of the locus is not perfect (e.g., D13S197). Multilocus genotype probabilities at these microsatellite loci do not show departure from the independence rule, unless the loci are closely linked. The allele size distributions at these $(CA)_n$ loci do not follow a strict single-step stepwise-mutation model. However, this feature does not compromise the ability to detect population affinities, when these loci are used simultaneously. The microsatellite loci examined here are present and, with the exception of the locus D13S197, are polymorphic in the chimpanzees, showing an overlapping distribution of allele sizes with those observed in human populations.

Introduction

Length polymorphism associated with tandem-repeat variation of dinucleotide $(dC\text{-}dA)_n \cdot (dG\text{-}dT)_n$ sequences—henceforth designated "$(CA)_n$ repeats"—in the human genome was first demonstrated in 1989, independently, by two groups of investigators (Litt and Luty 1989; Weber and May 1989). Since then, thousands of such $(CA)_n$ loci have been characterized. These loci are used extensively as gene-mapping markers because they are highly polymorphic and are widely and uniformly dispersed throughout the human genome. The most notable example of their application is the recent generation of high-density linkage maps of the human and mouse genomes (Dietrich et al 1994; Gyapay et al. 1994). These studies have demonstrated that a great majority of the loci that have been well characterized have heterozygosity levels in Caucasians >70%. However, few studies, so far, have attempted to characterize the population-genetic properties of this class of highly polymorphic loci. Furthermore, Caucasian samples of mixed origins have been analyzed principally to estimate allele frequency distributions at $(CA)_n$ loci. The few studies (Kamino et al. 1993; Bowcock et al. 1994; Di Rienzo et al. 1994) that surveyed these loci in non-Caucasian populations either involved too few individuals or considered amalgamated samples of individuals to represent a population. Therefore, accurate interpretation of population dynamics of $(CA)_n$ loci has been somewhat problematic.

In view of the considerations discussed above, we have characterized a set of eight $(CA)_n$ repeat loci located on human chromosome 13q in eight well-defined human populations encompassing a wide ethnic and geographic diversity. Furthermore, to study the antiquity of polymorphisms at these $(CA)_n$ arrays, a set of unrelated chimpanzees has been analyzed at the same loci, using primers designed from human sequence. The eight $(CA)_n$ repeat markers were intentionally chosen to examine how chromosomal linkage affects genotypic dependence between loci in unrelated individuals within populations. We have studied the extent of allele frequency variations at these loci and examined the conformity of genotype frequencies to their Hardy-Weinberg predictions and the extent of genotypic associations among loci. We have addressed two additional questions: (1) Does the variation observed at the $(CA)_n$ repeat markers provide any insight into the mechanism(s) of production of new alleles at these loci? (2) How useful is this class of polymorphism for studying human microdifferentiation?

Received July 20, 1994; accepted for publication November 16, 1994.

Address for reprints and correspondence: Dr. Ranjan Deka, Department of Human Genetics, University of Pittsburgh, A 300 Crabtree Hall, 130 DeSoto Street, Pittsburgh, PA 15261

* Current address: Department of Genetics, Stanford University, Stanford.

0002-9297/95/5602-0016$02.00

Am. J. Hum. Genet. 56:461–474, 1995

Populations and Methods

Populations

The Samoan (SA) sample represents a distinct Polynesian population, drawn from villages distributed throughout American Samoa and Western Samoa. The Dogrib Indian (DG) sample was drawn from Northwest Territories of Canada and represents the Na-Dene group. The Pehuenche Indians (PH), drawn from the Bio-Bio province of southern Chile, constitute a branch of Araucanian Indians. The New Guineans (NG) represent two linguistically different (Kalam and Gainj) but culturally similar interbreeding groups from the northern fringes of Papua New Guinea's central highlands. The Kachari (KA) are a distinct Mongoloid population living on the plains of the northeastern Indian state of Assam and are speakers of a Tibetoburman language. The Caucasian sample is represented by a German (GR) population drawn from northern Germany and by the unrelated parents from the CEPH (CP) cohort. The African sample is represented by the Sokoto (SO) population from northern Nigeria, who are predominantly members of the Hausa tribe. Detailed anthropological characteristics of several of these populations are presented elsewhere (Szathmary et al. 1983; Long et al. 1986; Deka et al. 1991). The Chimpanzee (CH) DNA samples were obtained from animals maintained at the Yerkes Regional Primate Research Center, Atlanta, and the Veterinary Resources Division, University of Texas M. D. Anderson Cancer Center, Bastrop, TX. All chimpanzees are African born and are presumably unrelated.

Laboratory Analysis

A summary of the genetic loci, their chromosomal locations, and primer sequences is given in table 1. For amplification of the $(CA)_n$ repeat loci, one of the primers was end-labeled using [γ-^{32}P]ATP and polynucleotide kinase T4. The amplified products were separated on 6% denaturing polyacrylamide gels. Following electrophoresis, the gels were dried, and allelic fragments were visualized by autoradiography. In addition to using an M13 sequence ladder on each gel as a size standard, the alleles were scored relative to genotypes determined in two individuals of the CEPH panel (see table 1). Figure 1 shows resolution of alleles at locus D13S71, using this protocol.

Data Analysis

Since all loci are autosomal (on chromosome 13) and detect codominant alleles, allele frequencies were obtained by the gene-counting method (Li 1976*a*). Such counting methods also readily yielded the allele-sharing statistics between populations.

Tests for Hardy-Weinberg expectations (HWE) are based on three test criteria: χ^2 test on the basis of contrasts of observed and expected heterozygosity/homozygosity; log-likelihood ratio statistic (Weir 1991); and Guo and Thompson's (1992) exact test for each locus-population combination. For each test, the levels of significance were empirically determined by shuffling (permutation) of alleles across individuals, as employed in our earlier studies (Chakraborty et al. 1991; Deka et al. 1991, 1992; Edwards et al. 1992).

Tests for genotypic independence across loci were done by a procedure described by Risch and Devlin (1992) and Morton et al. (1993), with the exception of the significance of the 2×2 contingency χ^2 statistic of differences of observed and expected match frequencies of genotype pairs of loci, which was judged by allele permutations as employed for the HWE tests.

Shriver et al.'s (1993) algorithm was used to generate a simulation database for mutation-model fitting. The simulation algorithm, described in that work, was extended to encompass a larger range of heterozygosity and a larger number of replications (100) of independent population histories of evolution so that errors due to resampling from the same replicated populations are minimized. For the infinite-allele model (IAM), the predictions for expected number of alleles, as well as probabilities of observing less than or equal to a given number of alleles, were analytically evaluated by following the theory described by Chakraborty and Weiss (1991), which is also a part of Shriver et al.'s (1993) algorithm.

Genetic-distance evaluations were made by employing the bias-corrected procedure for Nei's standard distance (D_S; Nei 1972) and modified Cavalli-Sforza distance (D_A; Nei et al. 1983). The standard errors of D_S were calculated by procedures described by Nei (1972). An analogous formula for standard errors of D_A is not available. For dendrogram construction from both genetic distances, we used Saitou and Nei's (1987) neighbor-joining method, in which the significance of branch lengths was evaluated by bootstrapping.

Results

The allele frequencies at the eight loci examined in nine populations, including the chimpanzees, are presented in the appendix, table A1. Although a comprehensive presentation of all data is not feasible, a few salient observations emerge from the allele frequency distributions. The number of alleles observed at these loci varies from 10, at the D13S124 locus, to 31, at the D13S197 locus. The spectrum of allelic variation is quite broad. For example, at the D13S71 locus, the allele frequency variations across populations have a complete overlap of allele sizes. Even the human and chimpanzee differences are reflected only at the level of allele frequency variation. In contrast, the D13S197 locus shows substantial variation, even at the level of allele sizes among human populations. The range of allele sizes is the largest in the Caucasian samples (GR and unrelated CEPH parents) and smallest in the PH sample. The chimpanzees have only two alleles, with frequencies of .98 and .02 at this locus.

Table 1

Summary of the Eight Microsatellite Loci Studied

Locus (Clone Name)	Chromosomal Location	Primer Sequence (5′ to 3′)	CEPH Reference Genotype[a] 133101	133102
FLT1	13q12	TTTGGCCGACAGTGGTGTAA AGGACCAAACCATGTCTGTC	170/182	168/168
D13S118 (Utsw1312)	13q14	CCACAGACATCAGAGTCCTT GAAATAGTATTTGGACCTGGG	190/194	190/190
D13S121 (Utsw1305)	13q31	GCTTGAGGTCTCTATGGAAA TTTCAGAACTCTGTACCAGGA	168/170	162/170
D13S71 (mfd44)	13q32-q33	GTATTTTTGGTATGCTTGTGC CTATTTTGGAATATATGTGCCT	75/75	75/75
D13S122 (Utsw1334)	13q31-q32	TGGAAACCACCACTCTACTT TGTGAACCTAGACTGGAATAAA	87/97	87/107
D13S197 (HKCA1)	13q31-q32	TTAATTCCCTGGAGCAGACG TCAGAGAAGTGGGCATGATG	97/97	126/128
D13S193 (HKCA5)	13q31-q32	GCAAGACCCCCATCTCTTAA CTCACCCCACTCCATGTTC	147/147	145/147
D13S124 (Mfd179)	13q21	CAAATTCAAATTCTTCCAGC ACTGTACTCCTGCATGTTAG	185/191	185/185

[a] Genotypes for two CEPH individuals, 133101 and 133102, used as reference markers.

Notwithstanding the fact that there is substantial variation in frequencies of alleles within each locus, alleles shared by all human populations account for most of the alleles. From the data presented in the appendix (table A1), it is clear that in 58 (90.6%) of a total 64 locus-population combinations, the combined frequency of alleles that are present in all the populations is ≥50%. This figure rises to ≥80% in 33 (51.5%) of the locus population combinations. The average proportion of alleles shared by all populations over all loci varies from 60.5% to 92.6%. The largest figures are observed in the DG (83.6%) and the PH (92.6%) populations, which also indicate that within-locus allelic variability is smaller in populations of smaller effective sizes. On the other hand, the SO population shows the lowest average (60.5%) among all the human populations studied. This population has several high-frequency alleles that are not shared by other populations.

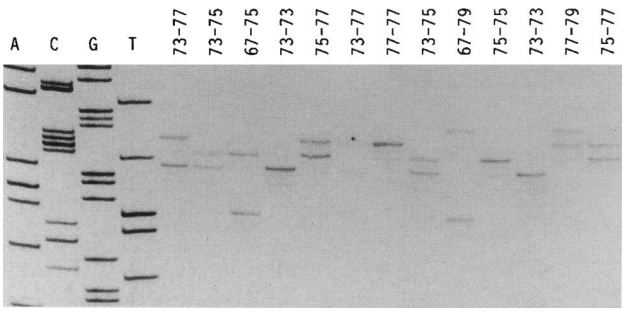

Figure 1 Resolution of PCR-amplified alleles at the D13S71 locus. The four left-hand lanes are an M13 sequencing ladder used as a size marker.

Table 2 shows the heterozygosity levels over all loci, as well as the average heterozygosity per locus (with their respective standard errors), in the examined populations. In general, the isolated (and presumably small) populations—for example, the PH, DG, and NG—have reduced average heterozygosity at the $(CA)_n$ loci. Interestingly, these populations also have a larger interlocus variability of heterozygosity levels. The smallest average heterozygosity, 48%, was observed in the PH population, with a range over all loci of 11%–75%. The CH sample has an average heterozygosity of 59%, with a range of 4%–88%. On the other hand, sample with larger effective sizes—for example, the SO, the two Caucasian populations, and the KA of northeastern India—have larger average heterozygosities, accompanied by a smaller interlocus variability. The largest average heterozygosity, 79%, is observed in the SO sample with a range of variability between 71% and 86%. These results are consistent with the hypothesis that a small average heterozygosity, together with the large interlocus variability, is indicative of small effective population size, as is evident in the samples of the CH, the PH, the DG, and the NG.

The results of the tests for conformity to HWE are shown in table 3 (for a description of the tests, see Populations and Methods), in which only the empirical levels of

Table 2

Observed and Expected Heterozygosities (%) at Eight $(CA)_n$ Loci

Locus	SA	DG	PH	NG	KA	GR	CP	SO	CH
FLT1:									
Observed	29.7 ± 4.4	21.5 ± 5.0	11.1 ± 3.0	46.9 ± 4.1	58.8 ± 6.8	28.4 ± 4.7	20.8 ± 4.6	79.5 ± 3.7	28.4 ± 4.7
Expected	31.9	20.0	10.7	44.2	61.6	29.5	20.7	80.4	72.8*
D13S118:									
Observed	52.8 ± 4.8	78.5 ± 5.1	64.4 ± 4.6	42.8 ± 4.0	76.5 ± 6.2	78.1 ± 4.6	67.5 ± 5.1	77.2 ± 4.2	69.1 ± 4.8
Expected	54.8	78.2	67.4	42.3	72.7	72.6	72.2	72.7	73.4
D13S121:									
Observed	57.7 ± 4.7	48.5 ± 6.2	35.8 ± 4.6	52.0 ± 4.0	77.6 ± 6.3	71.9 ± 4.6	78.2 ± 4.8	81.1 ± 3.5	80.8 ± 3.7
Expected	57.0	48.7	36.7	53.4	73.3	71.9	76.7	83.9	87.7
D13S71:									
Observed	76.6 ± 4.2	41.2 ± 5.9	24.7 ± 5.4	57.2 ± 3.9	84.9 ± 6.2	73.5 ± 4.5	73.9 ± 5.3	77.3 ± 4.1	20.4 ± 4.0
Expected	73.4	39.4	50.6*	58.0	71.1**	73.4	74.2	75.9	18.5
D13S122:									
Observed	87.0 ± 3.2	74.6 ± 5.7	63.6 ± 4.6	84.8 ± 3.0	78.0 ± 6.0	79.1 ± 4.2	87.0 ± 4.3	91.2 ± 3.2	57.3 ± 5.1
Expected	87.5	71.2	62.9	85.0	76.7	80.4	83.3	86.2	62.4
D13S197:									
Observed	73.2 ± 4.4	66.7 ± 5.2	56.4 ± 4.7	73.4 ± 3.6	64.7 ± 6.2	71.3 ± 4.3	77.9 ± 3.8	76.9 ± 3.6	0.0 ± 2.6
Expected	71.3	77.1*	54.5	77.1	73.1	78.0	87.4*	83.6*	3.7*
D13S193:									
Observed	84.7 ± 3.9	83.3 ± 5.2	75.7 ± 4.2	60.8 ± 4.0	74.4 ± 7.1	74.5 ± 4.4	71.1 ± 5.0	84.0 ± 4.0	70.3 ± 4.7
Expected	78.5	76.9	74.6	63.5	72.6	76.8	74.0	78.7	71.9
D13S124:									
Observed	67.0 ± 4.6	19.1 ± 5.0	29.8 ± 4.9	54.1 ± 3.9	56.6 ± 6.8	63.3 ± 4.9	59.0 ± 5.3	74.8 ± 3.9	80.9 ± 4.1
Expected	64.8	21.4	27.3	55.0	58.8	60.3	66.9	70.7	81.9
Average:									
Observed	66.0 ± 1.6	53.9 ± 2.2	46.1 ± 1.7	58.7 ± 1.4	71.3 ± 2.3	67.5 ± 1.7	66.8 ± 1.9	80.1 ± 1.4	52.9 ± 1.9
Expected	64.8	53.8	48.5	59.5	69.8	67.8	69.3	78.8	61.2

* $P \leq .05$.

** $P \leq .05$; observed heterozygosity > expected heterozygosity.

Table 3

Levels of Significance of Departure from HWE by Two Test Procedures—χ^2 Analysis of Observed and Expected Number of Heterozygotes and the Exact Test (E)

Locus	Test	SA	DG	PH	NG	KA	GR	CP	SO
FLT1	χ^2	.51	.54	1.00	.40	.72	.55	100	.77
	E	.65	1.00	1.00	.23	.14	.56	.63	.37
D13S118	χ^2	.55	1.00	.47	.88	.59	.21	.33	.21
	E	.01*	.58	<.01*	.93	.49	.74	.84	.01*
D13S121	χ^2	.90	1.00	.84	.58	.49	1.00	.75	.39
	E	.48	.54	.26	.32	.09	.32	.77	.14
D13S71	χ^2	.41	.65	<.01*	.82	.03*	1.00	1.00	.69
	E	.70	.22	<.01*	.58	.12	.10	.25	.76
D13S122	χ^2	.87	.55	.90	1.00	.86	.77	.39	.12
	E	.54	.69	.01*	.35	.75	.20	.54	.38
D13S197	χ^2	.73	.04*	.63	.28	.16	.09	.01*	.05*
	E	.02*	<.01*	.18	.72	<.01*	.01*	<.01*	.32
D13S193	χ^2	.13	.23	.81	.46	.85	.57	.52	.16
	E	.79	.23	.79	.91	.30	.05*	.11	.25
D13S124	χ^2	.67	.19	.44	.80	.74	.57	.11	.23
	E	.68	.23	.73	.88	.93	.51	.19	.01*

* $P = .05$. The empirical levels of significance are based on 2,000 replications of allele shuffling.

Table 4

Levels of Intra- and Interpopulation Variation at Eight $(CA)_n$ Loci in Eight Human Populations

Locus	F_{IS}	F_{ST}	H
FLT1	.0105	.1063	.3854
D13S118	−.0174	.0860	.6403
D13S121	−.0015	.0654	.5954
D13S71	.0048	.1381	.6388
D13S122	−.0132	.1211	.7818
D13S197	.0526	.0809	.7282
D13S193	.0374	.0683	.7448
D13S124	−.0228	.1911	.5260
Total (G_{ST})	...	.1065	...

significance (on the basis of 2,000 replications of permutations for each locus-population combination) are presented, since the values of the statistics, by themselves, offer no interpretation. For brevity, the results are shown for two test procedures: the χ^2 test (which is based on contrasts of observed and expected levels of heterozygosity at the loci) and Guo and Thompson's (1992) exact test. The empirical significance levels of the log-likelihood test (Weir 1991) were nearly always similar to those of the exact test.

Several features of the basic results of HWE are noteworthy. For example, when each locus-population combination is treated individually, at a 5% level of significance, several populations show deviations from HWE. However, significant departures from HWE are *not* consistently observed at all of the loci in a single population, nor at any single locus in all the populations. Of the total (over both tests) of 17 significant ($P < .05$) deviations, 8 (47%) are contributed by the D13S197 locus, which has a distinctive repeat motif (described below). This leads to the question of whether the observed deviations from HWE could be explained by chance departure due to multiple testing alone. For each population, the critical value, corresponding to the 5% level of significance, is ~0.64%, after Bonferroni correction of multiple testing (Weir 1991), since eight independent tests were conducted for each population for a particular test procedure. With correction for multiple testing, by excluding the D13S197 locus, departure from HWE is observed only in the PH population. Kinship computation (data not shown) indicates that this discrepancy is truly due to high levels of inbreeding in this population, which is probably due to small effective population size.

In addition, we have computed the bias-adjusted F_{IS} and F_{ST} levels (Nei 1987) and the average heterozygosity for the eight $(CA)_n$ repeat loci in the eight populations (table 4). The F_{IS}, or inbreeding coefficient, represents the extent of overall deviation from HWE. It is consistent with our results on the direct tests of HWE that the locus D13S197 shows the largest deviation from HWE, having 5.3% more homozygotes than expected. The F_{ST} can be understood as the proportion of the total variation that can be ascribed to differences between population allele frequencies. The F_{ST} for these loci ranged from 6.5%, at D13S121, to 19.1%, at D13S124. The average F_{ST} (G_{ST}) is 10.6% and is comparable to what has been reported, when traditional genetic markers in the human species have been used (Nei 1987).

Results of allelic association between loci, studied by pairwise independence of genotypic identities between individuals (a test developed by Risch and Devlin [1992] and Morton et al. [1993]; for description, see Populations and Methods), are shown in detail in the appendix (table A2). In all, among the 252 locus-pair/population–combination tests, 24 significant deviations from independence are observed. Of these, 13 occur with pairs of loci that are placed within 7 cM of each other, namely, D13S71, D13S122, D13S197, and D13S193 (Matise et al. 1994). The D13S197 locus is involved in 9 of 24 significant deviations. Samples from small isolated populations (CH, PH, DG, and NG) have accounted for 15 of these, as well. Like the tests of HWE, multiple testing was also involved in these tests. For each population, 28 locus-pair tests were performed. With Bonferroni correction, at the 5% level for individual tests, the adjusted critical level of significance would have been .0018. Examination of detailed data from the appendix (table A2) shows that, at this revised empirical level of significance, only three pairwise tests (D13S121-D13S122, D13S122-D13S197, and D13S122-D13S193, all in PH) are significant. In other words, in spite of syntenic location of these microsatellite loci, genotypic associations are detectable only when the loci are closely linked (in our case, within 7 cM of each other) and only in small isolated populations.

Since extensive diversity (high heterozygosity and a large number of alleles), conformity with HWE, and pairwise genotypic independence across these microsatellite loci have been shown, it is of interest to examine what maintains such polymorphisms and how new mutants arise at such loci. Examination of conformity of the number of alleles with their expectations based on gene diversity (heterozygosity) provides insight as to the probable mechanisms of mutations (the rationale and description of such tests are given in Populations and Methods). These results are shown in table 5, where the observed number of alleles and their expectations under the IAM and a single-step stepwise mutation model (SMM) are presented. In all, significant excess of allele numbers, in comparison with the single-step SMM predictions, are noted at 18 of the 72 locus-population combinations. Of these, only five have demonstrated significantly larger numbers of alleles, in comparison with the IAM predictions. The nine locus-population combinations that showed significantly fewer alleles in comparison with the IAM predictions are all within the 95% confidence limits of the SMM. In other words, of the 72 tests, 57 (79%) locus-population combinations of allele frequency distributions are in conformity with the IAM, while 54 (75%) are in conformity with the SMM. Only

Table 5

Fit of Neutral Mutation Models to Allele-Frequency Data

Population[a]		FLT1	D13S118	D13S121	D13S71	D13S122	D13S197	D13S193	D13S124
SA	1	4	6	6	7	12	11	6	4
	2	2.9 (.87)	5.6 (.70)	6.0 (.63)	10.4 (.14)	21.8[b] (.01)	9.6 (.77)	13.0[b] (.01)	7.5 (.09)
	3	2.8 (1.8, 4.2)	4.2 (2.5, 6.3)	4.5 (2.8, 6.8)	7.2 (4.5, 10.5)	14.5 (9.7, 20.6)	6.7[c] (4.2, 9.6)	8.8 (5.6, 12.6)	5.3 (3.3 ,7.7)
DG	1	3	7	6	8	6	7	6	3
	2	2.0 (.93)	11.4 (.08)	4.3 (.89)	3.4[c] (.99)	8.5 (.21)	10.9 (.11)	10.8[b] (.05)	2.1 (.91)
	3	2.2 (1.4, 3.4)	8.4 (5.1, 12.1)	3.7[c] (2.0, 5.2)	3.0[c] (1.9, 4.6)	6.4 (4.0, 9.3)	8.0 (4.9, 11.4)	7.9 (4.9, 11.3)	2.3 (1.5, 3.6)
PH	1	3	6	7	7	7	4	7	3
	2	1.5[c] (.98)	8.2 (.25)	3.3[c] (.99)	4.7 (.93)	7.2 (.58)	5.5 (.32)	10.9 (.11)	2.5 (.82)
	3	2.0[c] (1.0, 2.9)	5.9 (3.7, 8.7)	2.9[c] (1.9, 4.3)	3.8[c] (2.5, 5.6)	5.3 (3.2, 7.6)	4.3 (2.6, 6.2)	7.5 (5.1, 10.8)	2.4 (1.6, 3.9)
NG	1	6	6	8	5	11	10	7	3
	2	4.3 (.89)	4.1 (.91)	5.6 (.92)	6.5 (.34)	19.8[b] (.01)	12.8 (.23)	7.6 (.50)	5.9 (.11)
	3	3.5* (2.1, 5.2)	3.4[c] (2.1, 4.9)	4.2[c] (2.8, 6.1)	4.7 (2.9, 6.7)	12.3 (8.2, 17.5)	8.4 (5.5, 12.0)	5.3 (3.4, 7.7)	4.4 (2.8, 6.3)
KA	1	8	7	7	7	11	12	9	5
	2	5.9 (.89)	8.6 (.34)	8.7 (.33)	8.2 (.41)	10.9 (.73)	8.7 (.93)	8.0 (.75)	5.5 (.53)
	3	4.8* (2.9, 7.1)	6.7 (4.2, 9.8)	6.8 (4.2, 9.8)	6.4 (3.8, 8.9)	7.7 (4.7, 11.3)	6.8[c] (4.2, 9.9)	6.6 (4.0, 9.5)	4.5 (2.6, 6.4)
GR	1	9	7	8	5	13	16	11	6
	2	2.7 (.99)	9.8 (.20)	9.6 (.36)	10.2[b] (.03)	13.7 (.49)	12.3 (.92)	11.7 (.49)	6.5 (.53)
	3	2.6[c] (1.4, 3.9)	6.9 (4.2, 9.7)	6.8 (4.1, 9.5)	7.1 (4.4, 10.2)	9.5 (6.0, 13.5)	8.5[c] (5.6, 11.8)	8.1 (5.3, 11.3)	4.8 (2.8, 6.9)
CP	1	7	8	8	5	12	22	10	6
	2	2.0[c] (.99)	9.3 (.40)	11.1 (.18)	9.7[b] (.04)	15.3 (.20)	19.6 (.78)	9.9 (.61)	7.6 (.33)
	3	2.2[c] (1.1, 3.7)	6.8 (4.2, 9.4)	8.0 (4.9, 11.3)	7.2 (4.6, 10.1)	10.7 (7.2, 15.0)	13.8[c] (9.4, 19.4)	7.2 (4.6, 10.2)	5.7 (3.4, 8.0)
SO	1	14	9	11	8	13	18	12	8
	2	14.5 (.52)	10.2 (.41)	17.4[b] (.04)	11.6 (.15)	20.1[b] (.04)	16.9 (.68)	13.0 (.45)	9.8 (.33)
	3	9.6[c] (6.3, 13.7)	7.0 (4.4, 10.1)	11.5 (7.4, 16.1)	7.9 (5.0, 11.3)	13.1 (8.6, 18.2)	11.3[c] (7.4, 15.9)	8.8 (5.7, 12.6)	6.7 (4.3, 9.6)
CH	1	10	7	14	3	5	2	14	8
	2	9.7 (.63)	9.9 (.19)	20.2 (.06)	1.9 (.93)	6.8 (.30)	1.1[c] (.99)	9.5 (.97)	14.7[b] (.02)
	3	6.9 (4.3, 10.0)	7.1 (4.3, 10.3)	14.2 (9.2, 20.3)	2.2 (1.1, 3.4)	5.0 (3.1, 7.3)	1.7 (.8, 2.6)	6.7[c] (4.4, 9.7)	10.2 (6.8, 14.4)

NOTE.—The expected number of alleles for both mutation models is based on locus-specific heterozygosities shown in table 3. The levels of significance (*P*) for the IAM are based on analytical distributions of number of alleles. (Chakraborty and Weiss 1992), while the 95% confidence limits for SMM are from a simulated database applicable to such a sample size (appendix table A1) and heterozygosities (table 2), as per the algorithm of Shriver et al. (1993).

[a] 1 = Observed no. of alleles, 2 = Expected no. of alleles under IAM (*P*); and 3 = Expected no. of alleles under SMM (95% confidence interval).

[b] Number observed was significantly smaller than expected.

[c] Number observed was significantly larger than expected.

Table 6

Bias-Corrected Estimates of Genetic Distances Between Populations from Eight Microsatellite Loci

	SA	DG	PH	NG	KA	GR	CP	SO	CH
SA		.245	.169	.166	.120	.121	.155	.187	.711
DG	.346 ± .170		.101	.282	.187	.190	.186	.226	.697
PH	.231 ± .119	.097 ± .057		.255	.123	.131	.151	.225	.712
NG	.181 ± .089	.355 ± .149	.313 ± .136		.213	.222	.231	.242	.658
KA	.166 ± .062	.196 ± .068	.106 ± .026	.276 ± .128		.092	.112	.150	.643
GR	.156 ± .068	.201 ± .082	.124 ± .044	.315 ± .137	.117 ± .025		.033	.168	.680
CP	.190 ± .074	.212 ± .081	.147 ± .054	.337 ± .133	.145 ± .026	.005 ± .008		.166	.667
SO	.170 ± .040	.295 ± .103	.281 ± .094	.279 ± .087	.169 ± .062	.238 ± .065	.259 ± .068		.539
CH	1.822 ± .501	1.741 ± .498	1.901 ± .529	1.555 ± .443	1.527 ± .493	1.865 ± .506	1.860 ± .497	1.334 ± .375	

NOTE.—D_A values appear above diagonal; D_S values appear below diagonal.

five locus-populations combinations do not satisfy any of these two mutation-model predictions (i.e., are significantly larger than both model predictions). These occur in the PH, GR, and CEPH parents, for the FLT1 locus; in PH patients, for D13S121; and in the DG patients, for D13S71. Allele frequency predictions for 44 of the 72 locus-population combinations satisfy predictions of both mutation models.

We have estimated genetic distance between the examined populations on the basis of eight $(CA)_n$ repeat loci. The results of the computations are shown in table 6, where bias-corrected estimates of Nei's standard genetic distance (D_S, below the diagonal; Nei 1972) and of modified Cavalli-Sforza distance (D_A, above the diagonal; Nei et al. 1983) are presented. With respect to both measures, the chimpanzees are the most distant from all of the human populations, though, in relative terms, the human-chimpanzee distance does not correspond to the evolutionary time of interspecies comparison, when calibrated against the distances between all human populations (Deka et al. 1994).

Neighbor-joining trees (Saitou and Nei 1987) constructed from these distances (fig. 2), rooted by using the CH population as an outgroup, show that the SO population is the furthest from all other human populations. The CEPH parent sample is the closest to the GR population (the distance probably is statistically insignificant, judged from the standard error computations of D_S in table 6). The two trees are consistent to a large degree, except for the relative clustering of the two populations from the Pacific region (SA and NG). Indeed, the node separating the NG population from the remaining populations (except the SO) in the D_S tree is not significant (which is reflected by the low bootstrap value). In spite of these observations, the position of the Caucasians (GR and CEPH parents) in these trees is not anthropologically convincing, possible reasons for which are discussed below.

Discussion

The analyses of the eight microsatellite $(CA)_n$ loci presented here have several distinctive features, in comparison with the few published reports in this area. For example, Bowcock et al. (1994) used a set of 30 $(CA)_n$ repeat loci to study the evolutionary relationships of 14 human populations. However, the sample sizes used in their study (148 individuals in total, giving ~10 individuals/population) are too small to provide reliable estimates of allele frequencies (Chakraborty 1992). Even if their evolutionary inferences are correct, data from such small samples cannot fully characterize the extent of polymorphism (in terms of number of alleles and/or allele sharing), because of the low power of discrimination of the predictions from the two mutation models (Jin 1994).

While Bowcock et al. (1994) did not address the issue of mutation mechanisms at such loci, Di Rienzo et al. (1994) used 10 microsatellite loci to suggest that several microsatellite loci may follow a multistep SMM. Their sample sizes (46 Sardinians, 46 Egyptians, and 25 Africans) were small, and, their African sample came from at least nine different countries. With such heterogenous samples, their observed deviation from a single-step SMM can possibly be explained by population substructuring, rather than the multistep mutation mechanisms that they have advocated. In contrast, our results on mutation-model fitting offer an interesting insight. Of the 72 locus-population combinations, 44 fit both mutation models (IAM and SMM). Statistical congruence with the IAM prediction, however, does not negate the possibility of a "multistep" mutation mechanism (Li 1976*b*; Chakraborty et al. 1980). Only five locus-population combinations show a significant excess number of alleles, in comparison with the predictions of both models. As mentioned earlier, the FLT1 locus is responsible for three of these departures (in PH and GR and in CEPH parents), while D13S121 in PH and D13S71 in DG constitute the other two discrepant cases. Our predictions from both mutation models are based on mutation-drift balance; however, an observed excess number of alleles may be caused by recent expansion of population size. We believe that this probably is not the most likely cause. Closer examination of the allele size data (appendix table A1) shows that, in spite of the fact that all loci have

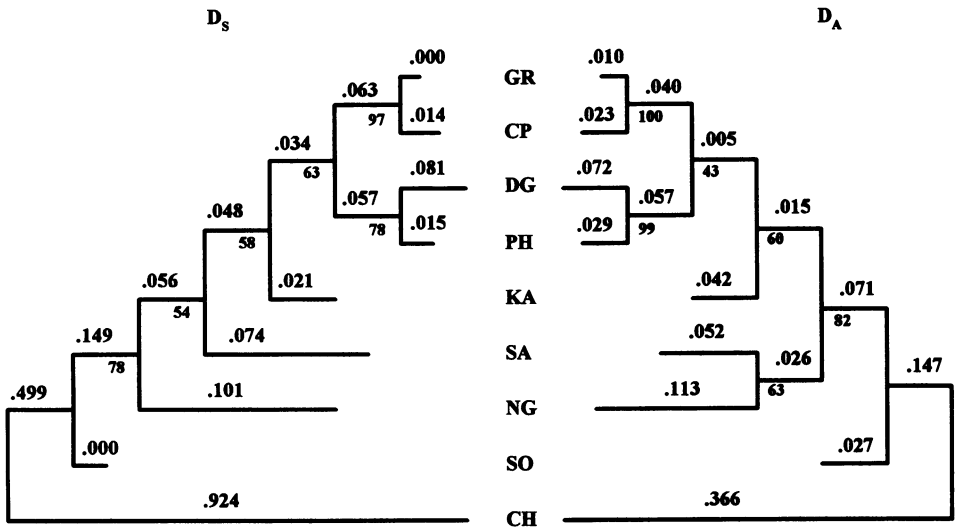

Figure 2 Neighbor-joining trees for the examined populations, based on D_A and D_S values. Branch lengths are not to scale. Bootstrap values, indicating the degree of support for each branch point, are shown below the line, as the percent of all replicates consistent with each branch point.

been labeled as $(CA)_n$ repeats, allele size alterations at such loci may involve insertion/deletion of single nucleotides or other, more-complex phenomena. For example, at the FLT1 locus in several populations (e.g., GR and CEPH parents) and at the D13S71 locus in DG and PH, the allele sizes are not always in increments of 2 bp, suggesting the existence of one or more 1-bp insertion/deletion polymorphisms. In addition, the D13S197 locus has a complex motif, where the CA repeat sequences are interrupted by an array of GC repeats. Hong et al. (1993) sequenced a presumed 25-repeat allele at this locus and detected a motif of $(CA)_5(GC)_8(CA)_{12}$. In other words, such imperfect motifs can cause disruption of allele size ladders, even if a molecular mechanism such as replication slippage is the predominant mechanism for the generation of new alleles. For such loci, dissection of alleles of different sizes would be needed to provide a better insight of the mutation mechanism, because it has been shown that microsatellite loci that are adjacent or close to each other show an excess of alleles (haplotypes) when the fine structure of the alleles at individual loci is ignored (Pena et al. 1994). Furthermore, three of the five discrepant locus-population combinations occur in two of the isolated populations (DG and PH), where small effective size could have caused the deviation. Recall that evidence of inbreeding due to limited effective size of the PH population is also found in HWE tests (table 3).

This study, as well as those of Bowcock et al. (1994) and Di Rienzo et al. (1994), addresses the evolutionary relationships among populations, using microsatellite loci, in terms of genetic distances and/or Wright's F_{ST} index. Results shown in figure 2, in conjunction with the general conclusions of Bowcock et al. (1994), indicate that the microsatellite loci are not only useful for such evolutionary studies but do indeed provide resolution beyond the power of traditional blood-group and protein loci.

In summary, we conclude that the $(CA)_n$ repeat loci have a greater diversity of allele frequencies across populations, in comparison with the minisatellite loci that are used in forensics (Budowle et al. 1991). The SO population has the largest degree of polymorphism, both in terms of number of alleles and in terms of locus heterozygosity. Conformity with HWE is generally observed, unless (*a*) a population is isolated and/or has a small effective size or (*b*) the involved locus has an imperfect or complex repeat motif. Hypervariability at microsatellite loci appears to produce a considerable degree of genotypic independence across loci, unless the loci are closely linked. Indeed, in a total sample of >800 individuals included in this study, we did not find any duplicate eight-locus genotypes, in spite of the fact that all loci are syntenic. A strict single-step SMM model of mutations may not be appropriate for all $(CA)_n$ repeat loci, as evidenced by 25% (18 of 72) of the locus-population combinations examined here. Recall that the IAM predictions can be regarded as approximations of a multistep SMM when the average number of allelic step changes by a single mutation is large (Chakraborty and Nei 1982). Genetic divergence between populations is also adequately reflected by the allele frequency differences between populations at such loci. This is consistent with other findings, as well (Bowcock et al. 1994; Deka et al. 1994).

Acknowledgments

We thank Dr. E. J. E. Szathmary for providing the Dogrib samples; Dr. G. Flatz for the German samples; Drs. P. E. Smouse, J. W. Wood, and J. C. Long for the New Guinea samples; Dr. S. T. McGarvey for the Samoan samples; Dr. F. Rothhammer for the Pehuenche Indian samples; Dr. F. I. Okoro for the Sokoto samples; and Dr. P. L. Alford for the chimpanzee samples. We thank Dr. Yixi Zhong for programming. This work was sup-

ported in part by grants GM 45861 (to R.D.), GM 41399 (to R.C.), and HG 00094 (to M.D.S.) from the National Institutes of Health, grant 92-IJ-CX-K024 (to R.C.) from the National Institutes of Justice, and support from the W. M. Keck Center for Advanced Training in Computational Biology at the University of Pittsburgh and the Carnegie Mellon University. The Samoan samples were collected through support by NIH grant AG 09375 to Dr. S. T. McGarvey.

Appendix

Table A1

Allele Frequencies (×1,000) at Eight (CA) Repeat Loci

Locus and Allele	SA	DG	PH	NG	KA	GR	CP	SO	CH
FLT1:									
156	0	0	0	0	0	0	0	4	102
158	0	0	0	0	0	0	0	0	45
160	0	0	0	0	0	0	0	0	28
164	0	0	0	0	0	0	0	0	28
166	0	0	0	7	0	32	6	26	0
167	0	0	0	0	0	5	6	0	0
168	806	892	944	710	559	837	890	389	0
170	9	0	0	28	59	37	32	43	23
172	0	54	0	0	0	16	0	60	11
174	0	0	32	0	0	0	0	90	318
176	0	0	0	0	0	0	0	4	398
178	0	0	0	0	49	0	6	34	40
180	5	54	0	17	10	0	0	47	0
182	180	0	23	231	264	58	45	90	6
184	0	0	0	7	29	5	0	132	0
186	0	0	0	0	0	5	13	73	0
188	0	0	0	0	0	5	0	4	0
190	0	0	0	0	10	0	0	4	0
200	0	0	0	0	20	0	0	0	0
Chromosome Data									
No. of chromosomes	222	130	216	290	102	190	154	234	176
Locus and Allele									
D13S118:									
176	0	0	0	0	0	0	0	0	6
180	0	0	0	0	0	0	0	0	42
182	0	0	0	0	0	0	0	0	256
184	0	0	0	0	0	0	0	48	65
186	0	0	0	89	0	16	39	35	393
188	0	69	29	0	157	104	130	158	208
190	648	269	505	750	441	224	234	469	30
192	5	46	10	49	10	0	19	35	0
194	134	300	117	26	167	427	442	39	0
196	88	92	107	0	10	31	13	35	0
198	83	215	233	69	176	182	117	149	0
200	42	8	0	16	39	16	6	31	0
Chromosome Data									
No. of Chromosomes	216	130	206	304	102	192	154	228	168

(*continued*)

Table A1 (continued)

Locus and Allele	SA	DG	PH	NG	KA	GR	CP	SO	CH
D13S121:									
150	0	0	0	0	0	0	0	0	19
154	0	0	0	0	0	0	0	0	16
156	0	0	0	0	0	0	0	0	13
158	0	0	9	0	0	0	0	0	96
160	0	0	0	0	0	0	0	108	38
162	0	0	0	0	71	73	83	36	26
164	0	53	5	3	10	36	32	95	58
166	617	697	789	668	429	495	429	320	160
168	117	23	18	56	112	88	103	67	179
170	189	68	92	66	255	52	71	58	199
172	27	7	41	26	41	73	58	149	103
174	18	152	46	30	82	115	141	86	64
176	32	0	0	105	0	68	83	36	32
178	0	0	0	46	0	0	0	9	0
180	0	0	0	0	0	0	0	36	6
Chromosome Data									
No. of Chromosomes	222	132	218	304	98	192	156	222	156
Locus and Allele									
D13S71:									
67	194	0	6	91	123	87	94	36	0
69	5	0	6	0	0	0	0	105	5
70	0	7	0	0	0	0	0	0	0
71	0	37	6	0	9	0	0	59	898
72	0	7	0	0	0	0	0	0	0
73	423	59	241	616	94	337	275	195	97
74	0	0	18	0	0	0	0	0	0
75	158	772	659	154	377	326	370	414	0
77	162	7	65	66	358	199	196	132	0
79	36	22	0	72	19	51	65	45	0
81	22	88	0	0	19	0	0	14	0
Chromosome Data									
No. of Chromosomes	222	136	170	318	106	196	138	220	186
Locus and Allele									
D13S122:									
75	0	0	0	0	0	0	0	0	34
77	0	0	0	0	0	0	0	0	320
79	0	0	0	0	0	0	0	0	517
81	0	0	0	0	0	0	0	0	67
83	0	0	0	10	0	11	33	4	62
85	0	0	0	0	10	0	0	0	0
87	120	0	36	0	140	154	123	40	0
89	0	0	4	76	10	0	0	27	0
91	56	0	0	0	70	49	46	22	0
93	88	0	0	41	110	143	91	159	0
95	227	111	564	90	440	368	351	62	0
81	0	0	0	0	0	0	0	0	67

(*continued*)

Table A1 (continued)

Locus and Allele	SA	DG	PH	NG	KA	GR	CP	SO	CH
D13S122: (*continued*)									
83	0	0	0	10	0	11	33	4	62
85	0	0	0	0	10	0	0	0	0
87	120	0	36	0	140	154	123	40	0
89	0	0	4	76	10	0	0	27	0
91	56	0	0	0	70	49	46	22	0
93	88	0	0	41	110	143	91	159	0
95	227	111	564	90	440	368	351	62	0
97	111	286	114	0	20	22	71	66	0
99	32	0	0	0	0	22	39	142	0
101	28	119	191	79	50	44	46	217	0
103	144	429	73	107	60	115	91	186	0
105	120	16	0	255	0	44	71	62	0
107	51	40	18	207	40	11	32	9	0
109	14	0	0	103	50	11	6	0	0
111	9	0	0	24	0	6	0	4	0
113	0	0	0	7	0	0	0	0	0
Chromosome Data									
No. of Chromosomes	216	126	220	290	100	182	154	226	178
Locus and Allele									
D13S197:									
87	0	0	0	0	0	5	0	0	0
97	0	23	0	65	10	117	163	0	0
98	0	0	0	0	0	0	6	0	0
99	0	0	0	0	0	5	13	0	0
101	0	0	0	0	10	0	0	0	0
105	0	0	0	0	0	0	0	0	981
109	0	0	0	0	0	0	0	0	19
112	0	0	0	0	0	5	0	0	0
118	0	0	0	0	0	5	6	32	0
119	0	0	0	0	0	5	0	0	0
120	9	0	0	3	0	0	6	83	0
121	0	76	0	0	20	32	13	14	0
122	391	273	614	3	412	319	234	278	0
123	0	83	0	0	0	0	91	19	0
124	353	356	259	144	294	319	175	250	0
125	0	61	0	0	0	0	52	9	0
126	0	129	118	47	127	27	58	120	0
127	5	0	0	0	0	0	6	32	0
128	5	0	0	3	20	37	52	28	0
129	0	0	0	0	10	0	13	0	0
130	60	0	0	3	0	16	6	5	0
131	23	0	0	0	39	0	0	37	0
132	93	0	9	119	29	27	26	28	0
133	19	0	0	0	10	11	6	19	0
134	42	0	0	349	20	21	13	23	0
135	0	0	0	0	0	0	0	14	0
136	0	0	0	263	0	0	6	5	0
138	0	0	0	0	0	0	39	0	0
139	0	0	0	0	0	48	6	0	0
142	0	0	0	0	0	0	6	0	0
145	0	0	0	0	0	0	0	5	0

(*continued*)

Table A1 (continued)

Locus and Allele	SA	DG	PH	NG	KA	GR	CP	SO	CH
Chromosome Data									
No. of Chromosomes	216	132	220	278	102	188	154	216	108
Locus and Allele									
D13S193:									
119	0	0	0	0	0	0	0	5	
123	0	0	0	0	0	0	0	0	5
125	0	0	0	0	0	0	0	0	16
127	0	0	0	0	0	0	0	5	11
129	0	0	9	0	13	11	46	80	5
131	315	265	238	143	218	170	151	236	0
133	257	326	173	546	423	133	132	354	352
134	0	0	0	0	0	0	0	0	5
135	0	61	0	52	26	0	7	127	385
137	0	0	0	0	0	0	0	28	88
139	0	0	0	0	0	11	7	9	27
141	0	0	0	0	13	21	26	5	11
143	54	0	5	28	13	37	13	0	22
145	104	189	164	213	51	64	53	24	0
146	0	0	0	0	0	11	0	0	0
147	99	144	378	7	231	409	454	108	60
149	171	0	33	10	13	122	111	14	5
151	0	15	0	0	0	11	0	9	0
Chromosome Data									
No. of Chromosomes	222	132	214	286	78	188	152	212	182
Locus and Allele									
D13S124:									
177	60	0	0	0	0	0	0	0	0
179	0	0	0	0	0	0	0	7	191
181	0	0	0	0	0	0	6	43	39
183	353	0	0	69	104	5	0	180	208
185	459	88	155	384	292	469	442	471	129
187	128	882	839	547	566	418	353	183	275
189	0	29	0	0	0	15	32	32	96
191	0	0	6	0	9	51	103	61	0
193	0	0	0	0	28	41	64	22	51
195	0	0	0	0	0	0	0	0	11
Chromosome Data									
No. of Chromosomes	218	136	168	318	106	196	156	278	178

Table A2

Probabilities for Test of Pairwise Independence of Loci

Pairs of Loci		SA	DG	PH	NG	KA	GR	CP	SO	CH
FLT1	D13S118	.75	.20	.58	.83	.24	.60	.08	.39	.46
	D13S121	.34	.47	.53	.67	.28	.41	.10	.56	.20
	D13S71	.24	.57	.85	.62	.23	.39	.55	.50	.37
	D13S122	.44	.93	.33	.60	.31	.13	.37	.79	.36
	D13S197	.97	.71	.67	.74	.54	.41	.91	.70	.09
	D13S193	.63	.72	.38	.41	.23	.64	.62	.06	.06
	D13S124	.96	.45	.18	.18	.28	.02*	.12	.54	.07
D13S118	D13S121	.23	.53	.92	.68	.02*	.51	.15	.92	.43
	D13S71	.79	.31	.05*	.63	.40	.84	.64	.23	1.00
	D13S122	.24	.16	.55	.04*	.02*	.35	.09	.79	.05*
	D13S197	.67	.54	.22	.68	.11	.26	.53	.44	.09
	D13S193	.53	.32	.16	.69	.29	.41	.52	1.00	.64
	D13S124	.86	.41	.92	.18	.20	.24	.02*	.39	.11
D13S121	D13S71	.67	.90	.43	.49	1.00	.96	.57	.55	.47
	D13S122	.02*	.92	<.01*	.48	.23	.95	.21	.87	1.00
	D13S197	.65	.33	.09	.56	.92	.36	.06	.87	.65
	D13S193	.77	.47	.60	.99	.24	.95	.10	.63	.70
	D13S124	.59	.99	.33	.94	.87	.14	.92	.85	.31
D13S71	D13S122	1.00	.38	.01*	.85	.82	.56	.21	.49	.35
	D13S197	.01*	.46	.04*	.02*	.60	.90	.33	.63	1.00
	D13S193	.01*	.01*	.43	.07	.26	.21	.90	.10	.65
	D13S124	.43	.97	.06	.46	.91	.74	.20	.26	.54
D13S122	D13S197	.38	.02*	<.01*	.04*	.08	.01*	.81	1.00	.05*
	D13S193	.77	.01*	<.01*	.12	.27	.59	.45	.81	.01*
	D13S124	1.00	.83	.90	.38	.92	.83	.32	.85	.76
D13S197	D13S193	.37	.22	.83	.12	.50	.75	.64	.40	.16
	D13S124	.29	.47	.69	.03*	.22	.21	.15	.01*	.25
D13S193	D13S124	.59	.51	.009	.32	1.00	.30	.93	.21	.32

References

Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL (1994) High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368:455–457

Budowle B, Giusti AM, Waye JS, Baechtel FS, Fourney RM, Adams DE, Presley LA, et al (1991) Fixed-bin analysis for statistical evaluation of continuous distributions of allelic data from VNTR loci, for use in forensic comparisons. Am J Hum Genet 48:841–855

Chakraborty R (1992) Sample size requirements for addressing the population genetic issues of forensic use of DNA typing. Hum Biol 64:141–159

Chakraborty R, Fornage M, Gueguen R, Boerwinkle E (1991) Population genetics of hypervariable loci: analysis of PCR based VNTR polymorphism within a population. In: Burke T, Dolff G, Jeffreys AJ, Wolff R (eds) DNA fingerprinting: approaches and applications. Birkhäuser, Basel, pp 127–134

Chakraborty R, Fuerst PA, Nei M (1980) Statistical studies on protein polymorphism in natural populations. III. Distribution of allele frequencies and the number of alleles per locus. Genetics 94:1039–1063

Chakraborty R, Nei M (1982) Genetic differentiation of quantitative characters between populations or species. Genet Res 39:303–314

Chakraborty R, Weiss KM (1991) Genetic variation of the mitochondrial DNA genome in American Indians is at mutation-drift equilibrium. Am J Phys Anthropol 86:497–506

Deka R, Chakraborty R, DeCroo S, Rothhammer F, Barton SA, Ferrell RE (1992) Characteristics of polymorphism at a VNTR locus 3′ to the apolipoprotein B gene in five human populations. Am J Hum Genet 51:1325–1333

Deka R, Chakraborty R, Ferrell RE (1991) A population genetic study of six VNTR loci in three ethnically defined populations. Genomics 11:83–92

Deka R, Shriver MD, Yu LM, Jin L, Aston CE, Chakraborty R, Ferrell RE (1994) Conservation of human chromosome 13 polymorphic microsatellite $(CA)_n$ repeats in chimpanzees. Genomics 22:226–230

Dietrich WF, Miller JC, Steen RG, Merchant M, Damron D, Nahf R, Gross A, et al (1994) A genetic map of the mouse with 4,006 simple sequence length polymorphisms. Nat Genet 7:220–245

Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB (1994) Mutational processes of simple-sequence repeat loci in human populations. Proc Natl Acad Sci USA 91:3166–3170

Edwards A, Hammond HA, Jin L, Caskey CT, Chakraborty R (1992) Genetic variation of five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics 12:241–253

Guo SW, Thompson EA (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 48:361–372

Gyapay G, Morissette J, Vignal A, Dib C, Fizames C, Millasseau P, Marc S, et al (1994) The 1993–94 Généthon human genetic linkage map. Nat Genet 7:246–339

Hong H-K, Giorda R, Trucco M, Chakravarti A (1993) Microsatellite repeat polymorphism at the D13S197 locus. Hum Mol Genet 2:337

Jin L (1994) Population genetics of VNTR loci and their applications in evolutionary studies. PhD thesis, University of Texas, Houston

Kamino K, Nakura J, Kihara K, Ye L, Nagano K, Ohta T, Jinno Y, et al (1993) Population variation in dinucleotide repeat polymorphism at the D8S360 locus. Hum Mol Genet 2:1751

Li CC (1976*a*) First course in population genetics. Boxwood, Pacific Grove, CA

Li W-H (1976*b*) A mixed model of mutation for electrophoretic identity of proteins within and between populations. Genetics 83:423–432

Litt M, Luty JA (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am J Hum Genet 44:397–401

Long JC, Naidu JM, Mohrenweiser HW, Gershowitz H, Johnson PL, Wood JW, Smouse PE (1986) Genetic characterization of Gainj- and Kalam-speaking peoples of Papua New Guinea. Am J Phys Anthropol 70:75–96

Matise TC, Perlin M, Chakravarti A (1994) Automated construction of genetic linkage maps using an expert system (MultiMap): a human genome linkage map. Nat Genet 6:384–390

Morton NE, Collins A, Balazs I (1993) Kinship bioassay on hypervariable loci in Blacks and Caucasians. Proc Natl Acad Sci USA 90:1892–1896

Nei M (1972) Genetic distance between populations. Am Nat 106:283–292

——— (1987) Molecular evolutionary genetics. Columbia University Press, New York

Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. J Mol Evol 19:153–170

Pena SDJ, De Souza KT, Andrade MD, Chakraborty R (1994) Allelic associations of two polymorphic microsatellites in intron 40 of the human von Willebrand factor gene. Proc Natl Acad Sci USA 91:723–727

Risch N, Devlin B (1992) On the probability of matching DNA fingerprints. Science 255:717–720

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Shriver MD, Jin L, Chakraborty R, Boerwinkle E (1993) VNTR allele frequency distributions under the stepwise mutation model: a computer simulation approach. Genetics 134:983–993

Szathmary EJE, Ferrell RE, Gershowitz H (1983) Genetic differentiation in Dogrib Indians: serum protein and erythrocyte enzyme variation. Am J Phys Anthropol 62:249–254

Weber JL, May PE (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am J Hum Genet 44:388–396

Weir B (1991) Genetic Data Analysis. Sinauer, Sunderland, MA